

Selecting Shelter Dogs for Service Dog Training

Emily Weiss

*Department of Psychology
Wichita State University and
Sedgwick County Zoo
Wichita, Kansas*

Service dogs are an essential aid to persons with disabilities, providing independence, mobility, and improved self-esteem. Because of these proven benefits, the growing use of service dogs is creating a demand and supply crisis. One major cause is the 50% average dropout rate for dogs selected for training. Weiss and Greenber (1997) recently found that a dog, successful on the most commonly used selection test items, was as likely to be either a poor or a good candidate for service work. The experiment presented here evaluated test items developed by the author in 15 years of experience with dogs. The test items were administered to 75 dogs from the Kansas Humane Society. Once tested, the dogs received obedience and retrieval training. The experiment assessed each dog on behavior over 5 weeks of training versus performance on each selection test item. A subset of the selection items, combined in a regression analysis, accounted for 36.4% of the variance with $R = 0.603$. This research also revealed a reliable test for dog aggression without risking injury to dog or tester. Items for testing included fear, motivation, and submission. Another set of selection items reliably predicted the trait of "high energy" commonly described as "high strung." Future research should involve investigating the effectiveness of both cortisol levels and blood pressure in predicting traits to help strengthen the predictive value of the tool and then testing on dogs trained to be full service dogs.

Dogs have been associated with humans for more than 14,000 years (Prestrude & O'Shea, 1998). They have served as companions, guards, hunters, herders, medical subjects, and guides. Selection of a dog for a particular task often is

based on matching the task to the breed, attempting to choose a dog with the appropriate physical and psychological attributes. This procedure is inexact and frequently results in potentially costly mismatches. Recently, the selection of dogs to be used to assist humans with physical or mental disabilities has been made even more difficult by the introduction of pound and animal shelter dogs as candidates. Chosen correctly, however, these dogs can decrease costs and dramatically increase the pool of dogs that can be trained for service work.

The use of dogs to increase independence is not a new idea. The first organized dog guide school in America, The Seeing Eye, was established in 1929 (Blasch & Stuckey, 1995). The use of walking canes at this time was much more prevalent, but many found that the use of the dog offered more freedom, and walking with a dog was less socially stigmatizing than using a cane.

More widespread use of guide dogs occurred at the end of World War II. In 1942, a group of women opened Guide Dogs for the Blind, a school with the goal of helping veterans blinded in the war obtain guide dogs (Pfaffenberger, Scott, Fuller, Ginsburg, & Bielfelt, 1976). At that time, the supply of dogs was quite low. The school was founded to provide trained animal companions and find better selection processes to increase the number of dogs available for those who needed them. The school became the first to develop a selection tool for service dogs.

The service dog field has grown and now uses dogs who can perform several different types of tasks. Each type of service requires a different type of dog, either in physical makeup or in psychological attributes. The guide dog should be large and possess a good attention span and high endurance. The guide dog should be very trainable, not territorial, and neither highly submissive nor dominant. Most often trained for these purposes was the German Shepherd, but use of that breed has declined, and use of Labrador Retrievers and Golden Retrievers has increased. Trainers switched to the retrievers because the German Shepherd has difficulty adapting to down time (lying passively by while the owner eats, works, sits in class), whereas the retrievers are much more adaptable to this lack of activity.

SELECTION OF SERVICE DOGS

In 1947, Pfaffenberger and Scott developed the first puppy test for the selection of guide dogs (Pfaffenberger et al., 1976) and used it at their school, The Guide Dogs for the Blind. Later, Scott and Fuller (1965) published studies—on which this test was based—on the genetics of the social behavior of the dog. They reported that the behavioral development of dogs could be divided into different stages based on social changes in the puppy. These stages are the neonatal, transitional, socialization, juvenile, pubertal, and parental.

During the socialization period (3 to 12 weeks), the puppy easily forms attachments. If not exposed to humans during this period, a puppy will react extremely

fearful when later exposed to humans (Scott & Fuller, 1965). Scott and Fuller's findings are the basis for the selection test they later developed at the Guide Dogs for the Blind school. Puppies are tested many times during the socialization period to determine if they are suitable for the training regimen.

Pfaffenberger et al. (1976) reported that the high rate of dogs entered in the training process but unable to complete it decreased to approximately 50% with the use of the selection test. The test became an important asset for training programs because, even though a large failure rate, 50% is less than the rate when no selection test was used. Results of a subsequent assessment of Pfaffenberger's test showed that 58.6% of the dogs entering training completed training. These results are similar to what was originally reported (Pfaffenberger et al., 1976). The failure rate in some programs can be misleading because some programs do not consider animals failures if they do not become guide dogs and shift them into the breeding program or a therapy dog program.

A series of studies by Goddard and Beilharz (1983, 1984, 1985a, 1985b, & 1986) examined the ability to predict adult behavior, particularly fear behavior, in puppies to be trained as guide dogs. Although excitement or inhibition are considered important behaviors, their studies focused most often on "fearfulness," the behavior they reported to correlate highest with failure to complete training. Their work, like that of Pfaffenberger et al. (1976), used puppies bred for guide dog use. They found that the most useful tests for assessment involve the puppy's reaction to a strange person, a strange dog, and certain unusual objects (Goddard & Beilharz, 1986). They found that when the puppies are 3 months of age there is some predictability of adult behavior and that this predictability increases with the age of the puppy (Goddard & Beilharz, 1984).

Others (Bartlett, 1979, 1985; Campbell, 1972) have designed puppy tests for companion dogs based on the evaluation of behavioral tendencies such as dominance and submission. However, empirical studies show that these tests have not proved to be good predictors of adult behavior (Beaudet, Chalifoux, & Dallaire, 1994; Young, 1988). It seems that there are so many variables that can affect development of traits that puppy behavior is not a good predictor of adult dog behavior.

Wilsson and Sundgren (1998) recently conducted a very large study focusing on the correlation between behavior of 8-week-old puppies and behavior at approximately 1 year of age. Using 630 German Shepherd dogs, they found that the puppy test was not a reliable predictor of adult suitability for service dog work.

Because of the high costs of breeding and the volunteer hours necessary, some in the service dog industry have recently turned to adult and sub-adult dogs in shelters and pounds as a ready source of inexpensive, trainable dogs. This has presented a new challenge to the selection process. These dogs cannot be chosen using the tests of Pfaffenberger et al. (1976) or Goddard and Beilharz (1984) because those tests were designed for use with puppies; dogs chosen from shelters are usually between 8 months and 2 years of age. Agency selection tests are based,

at least in part, on Pfaffenberger's test, although most centers use a combination of tasks including observations of the dog's behavior in the cage, the dog's reaction to pain (usually a toe pinch), and the dog's reaction to novel stimuli (e.g., an umbrella opening suddenly). Even with the use of a selection test, centers report between 40% to 80% failure rate. This failure rate was acceptable when compared to no selection test at all, but obviously a lower failure rate would result in great savings of money and time and lead to the increased availability of service dogs.

There are few studies of the predictability of behavior of dogs from a shelter or pound environment. One study looked at the predictability of problem behavior of shelter animals adopted as companion animals, reporting that their test predicted 74.7% of the future behavior problems that the tested dogs displayed (Van der Borg, Netto, & Planta, 1991). Although this is encouraging, many of the dogs used in the study would never have been considered for service dog use. Dogs with overt behavioral problems, such as territorial aggression or extreme fear, are passed over by service dog trainers without testing.

The older age and accompanying maturity of shelter dogs might mean that, with the correct test, behavior might be better predicted than in puppies bred specifically for service tasks. As indicated earlier, Goddard and Beilharz (1984) found that ability to predict adult fearfulness in a puppy increased as the age of the puppy increased. The greatest problem identified by service dog training centers is that the dog's overall behavior undergoes large changes, beginning approximately 3 to 4 weeks into the training process; specifically, problems with both fear and territorial aggression increase.

The problems faced by training centers in choosing dogs are related to the effectiveness of the selection tests being used. When the original puppy tests were first used, they improved the selection process for dogs to be used by people with physical disabilities. Since then, there have been no further improvements on these tests. Changes in the breeds used by training centers and changes in the population from which some dogs are chosen suggest a need to reassess the selection criteria.

ASSESSMENT OF CURRENT SELECTION TOOLS AND NEW SELECTION CRITERIA

An earlier project assessed the selection test most commonly used by training centers (Weiss & Greenberg, 1997). Given my long-term experience in the handling and training of dogs and based on discussions with officials of many of the leading service dog training centers, we hypothesized that the selection test is, by itself, a poor predictor of training success, and our results supported the hypothesis. Based on this early work and extensive literature reviews, experience, and interviews with service dog agencies, a new selection tool was developed. All components are based, on some level, on research from various fields of ani-

mal behavior. An open field test, here called Activity level, is a well-established emotionality-defining tool in the field of animal behavior and psychobiology (Weiss & Greenberg, 1998). A study of police dogs conducted by Martinek (1972) found a strong correlation between successful candidates and an average level of activity in an open field test. The experiment was designed to evaluate selection items that are successful at predicting good potential service dog behavior, trainability, pet quality, fearfulness, and other traits.

Design

Animals. Forty-six male and 29 female dogs between 6 months and 2 years were acquired from the Kansas Humane Society, Wichita, Kansas. Nineteen of the dogs were spayed or neutered. The breed types and numbers of each breed were 19 Labrador retrievers and Labrador mixes; 16 German Shepherd dogs and mixes; 12 Rottweilers and mixes; 5 border collies and mixes; 5 golden retrievers and mixes; 4 Australian Shepherd and mixes; 14 mixes that were not distinguishable as to any particular breed. All dogs were between 22 to 50 kg.

Once selected, the dogs were transported to Chisholm Creek Kennels in Wichita. They were housed in individual indoor and outdoor cages approximately 6 m × 1.2 m. Each dog was given a 5-day treatment of antibiotics, all vaccinations, heart worm preventive, administered Panucur (DPT Labs, San Antonio, CA) to remove internal parasites, and flea and tick control. Each dog was given chew toys and Kongs (Kong Co., Golden, CO) or Buster Cubes (Kruuse A/S, Denmark; enrichment devices) daily for play enrichment. All dogs were given access to a play yard for approximately one half hour 5 to 6 days a week.

Procedure. All four dog-selection testers (referred to as testers) had at least 5 years of professional dog experience. Two testers each had more than 16 years of experience. Each of the two dog trainers (referred to as trainers) had more than 13 years of professional animal training experience.

Procedure Conducted in Four Phases

Phase 1: Selection. This phase was conducted at the Kansas Humane Society. The testers made an initial walk through the adoption area to identify and tag the cages that housed animals of the appropriate breeds, sizes, and ages. Once the initial check was completed, dogs were selected randomly for testing from the tagged cages. All tagged dogs were used in the study.

The individual selection test items are described in Table 1. Each dog was administered all test items, but in a random order except where noted. All selection

TABLE 1
Selection Items Tested for Effectiveness in Predicting Future Behavior

Controlled walk	Spend up to 5 minutes working on a controlled walk, correcting the dog for pulling on the leash by using a collar correction and using high praise.
Pass	The dog should walk by the handler for the majority of the time, be friendly, and tolerate corrections.
Fail	Sustained lack of attention toward handler; no improvement in pulling or walking position during the 5 minutes.
Fetch	Take the dog to a small fenced in area (6.1 meters × 6.1 meters) and using a brightly colored soft rubber toy or tennis ball, get the dog's attention by swinging it back and forth in front of you. Once the dog is attending to the toy throw it in front of you and then get involved in the catch—clap your hands, smile and use a lot of energy. If the dog brings the item back toward you, try to engage the dog in a bit of pulling and then repeat the sequence once you have removed the toy from the dog's mouth.
Pass	Dog engages in any sort of game with handler. Ideally the dog retrieves the toy repeatedly.
Fail	Ignoring the handler; investigating area relentlessly throughout test.
The pinch	The test requires a person to pinch between the dog's toes and wait for a response, slowly pinching harder and harder until the dog responds.
Pass	Dog gently withdraws its paw; no response
Fail	Yelping; submissive urination; growling, or biting
Sensitivity	The handler should sit in an armless chair in a quiet area away from distractions and place the dog in front of him/her. Beginning at the haunches and working up towards the ears pinch the loose skin lightly all over the dog. The pinch should have some pressure, but should not cause pain.
Pass	Dog leans into the touch; dog solicits more interaction; dog accepts touch without moving.
Fail	Dog in constant movement; dog not interacting with handler; dog growls or bites.
The Stranger	While the handler is walking the dog, a stranger wearing a mask suddenly pops up from behind an object—or around a corner and walks toward the dog without looking at it.
Pass	Dog approaches stranger in calm and friendly manner; dog initially startles but quickly recovers and continues to walk.
Fail	Dog remains strongly focused on stranger; dog startles and does not recover after 10 seconds; dog growls or barks.
Umbrella test	In a closed room, drop the leash and with a closed umbrella stand in front of the dog and open it. Place the opened umbrella on the floor on its side.
Pass	Dog initially startles but quickly recovers and approaches umbrella.
Fail	Dog startles and does not recover in 10 seconds.
Sound sensitivity	While in an enclosed room, drop the leash and bang a spoon against a pan once, wait 5 seconds and then bang continuously every half second.
Pass	Dog approaches pan or handler in a calm manner; dog initially startles but quickly recovers.
Fail	Dog startles and does not recover in 10 seconds.
The stare	Kneel down in front of the dog's cage. While the animal is calm, look directly into its eyes without speaking. Continue for 30 seconds.
Pass	Dog stays in the front of the cage trying to interact with the handler while avoiding direct eye contact.
Fail	Constant direct eye contact; dog stays in the back of the kennel; dog does not solicit interaction.

- The object While the handler is walking the dog, they should come across an unusual object such as a large animal figure.
- Pass Dog seems to ignore object; dog initially startles but quickly recovers and walks on.
 - Fail Dog startles and does not recover.
- Other dog While the handler is walking the dog on a leash, have another dog (large and of the same sex as the dog being walked) tied on a long leash near by. The handler should walk the dog past the tied dog staying about 5 feet from it.
- Pass Dog ignores other dog; dog approaches in play position; dog approaches in a submissive posture.
 - Fail Dog approaches while barking and/or growling; dog is fixated on other dog, and the handler is unable to distract or continue walking the dog; dog approaches tail high and flagging, ears in forward posture and body stiff.
- Settle This test will always be the last test given. Gently but forcefully place the dog on its side and hold it down by bracing one hand on the neck, and the other on the haunches. Remain in this position for 40 seconds or if earlier, when the dog is still.
- Pass Dog calmly collapses into position.
 - Fail Dog struggles; dog submissively urinates; dog growls or attempts to bite.
- Activity level The tester will walk the dog into a empty room, quickly remove the leash and exist the area. Video taping should begin immediately after handler has exited. The dog's activity (amount of squares crossed, amount of vertical elevations or rearings, and the amount of sniffing) should be recorded during one 4 minute session. Squares crossed will be determined using a overlay on monitor during the review of the tape.
- Pass Average level of activity; little to no vertical activity.
 - Fail Excessively high or low activity; high vertical activity; panic behavior.
- Jumping at people A record will be kept of the number of times the dog jumps on anyone at any time during the testing sessions.
- Pass Dog does not jump on tester, or jumps on tester only one time.
 - Fail Repeated jumping on tester.
- Play aggression A record of any play aggression (i.e., barking at handler while wagging tail, play nips, etc.) at any time during the testing sessions will be kept.
- Pass Little to no play aggression
 - Fail More than one instance of play aggression
- The gut Once all items have been administered, the tester/handler, and the videotaper make a record of their "gut prediction" for the success of the dog they just tested. Analysis will determine if this measure is a reliable predictor of success.
-

items were videotaped for later analysis, and notes were taken after each item. All selectors involved in the selection have extensive professional experience with dogs. Because the study was designed to examine the effectiveness of the individual items on the selection test, all dogs tested were brought into the study, regardless of their performance in the test.

Phase 2: Obedience training. All dogs who were administered the selection tool, regardless of performance, were obedience trained. Obedience trainers were blind to the dogs' behavior during the selection process. Obedience training was conducted using the Volhard method (Volhard & Volhard, 1992), which involves positive reinforcement (high praise), negative reinforcement (ignoring the animal and chain collar corrections), and punishment (chain collar corrections). Training included basic commands required to teach the specific tasks: no, sit, stay, stand, come, down, heel, off, leave it, quiet. To simulate real life situations, training included testing the dogs in a shopping center.

Completion of the obedience phase was determined by the dog's ability to respond correctly 75% of the time to the first verbal command. This criterion is slightly higher than that used in other task studies because this phase has a lower level of learning difficulty compared to other studies.

Phase 3: Retrieval training. In this phase, dogs were trained to return a retrieval dumbbell. This was taught using the Volhard motivational retrieval method (Volhard & Volhard, 1992), which uses a positive reinforcer of food. The criterion for completion was 70% correct responses using one command and one correction (if necessary) in the familiar training environment.

Phase 4: Evaluation. Both obedience trainers assessed each dog using a 10-item questionnaire (see the Appendix). Each trainer filled out the questionnaire separately and then was reviewed by both trainers. Any discrepancies were discussed, and differences of one point were averaged. The interrater reliability was very high, with only one-point discrepancies between each trainer's ratings. This was the primary evaluation as not all dogs completed the retrieval task.

Furthermore, each animal was tested for the ability to retrieve an item in three situations likely to be encountered while working as a service dog. All retrievals occurred while the dogs were on leash. The three situations were (a) retrieving an item in a quiet area of the training center (home environment) with only the trainer present, (b) retrieving an item in a quiet area of the training center (home environment) with a male stranger present, and (c) retrieving an item inside the home environment while a loud noise (a bang against the wall) was occurring. The banging began immediately after the retrieval command was given. The number of corrections (if any) needed to have the dog complete the task was recorded.

To attempt to assess the motivation level of the dogs, a rating of the type of reinforcement needed to train and maintain a behavior was used. This was assessed via notes taken during training. Notes were specifically taken on the type of reinforcer needed to train new, and maintain old, behaviors.

Results

Training. Forty of the 75 dogs learned both the obedience and the retrieval tasks in the allotted 5 weeks of training. The time allotted for the obedience training and the retrieval training was determined first by the minimum amount of time needed for a dog's behavior to return to baseline after a move and stress, which has been reported from training centers across the country as 3 weeks. Five weeks was the minimum amount of time needed to train all tasks, and because of the cost of housing and training, 5 weeks was the maximum amount of time allotted to house dogs. Thirty-three dogs learned all the obedience tasks and had entered the retrieval portion of training but had not reached criterion in time. Two dogs did not complete training because it was determined that the trainers were at risk of injury because of their aggressive behavior.

The mean amount of time needed to complete obedience training was 207.58 min, (range 135 to 345). The mean time for learning the retrieve task was 217.10 min (range 92.00 to 345). The mean total training time was 481.72 min. Because of the experience differences of the dogs who enter the shelter, one would expect a large difference in the minimum and maximum amount of time to learn obedience, the large range of learning times for the retrieve was not expected, as none of the dogs knew a trained retrieve.

Post Training

Trainer's rating. On completion of training each group of dogs, the two trainers filled out an assessment form for each dog. Trainer ratings were based on a 6-point scale (see Table 2) ranging from 6 (*most desirable behavior*) to 1 (*least desirable behavior*). For example, in the category of good service (good service dog potential), 5 dogs were given the highest rating and 15 the lowest. Dogs also were rated on the degree of reinforcement needed to produce a consistent response when learning a new command. The categories and number of dogs who needed that level of reinforcement were as follows: verbal praise (13 dogs, 17%); food (12 dogs, 16%); food and praise (21 dogs, 28%); high-energy verbal encouragement praise (14 dogs, 18.6%); high level of physical patting (7 dogs, 9.33%); collar corrections (8 dogs, 10.5%).

TABLE 2
Results From Trainers' Assessment of Dogs

Category	Ratings of										
	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
Easy to train	2	1	5	2	3	6	12	11	8	10	15
Fearful	2	5	8	4	9	5	9	15	7	6	4
Dominant	1	0	5	4	4	8	13	11	12	12	5
Submissive	1	7	8	9	16	15	5	8	3	3	0
Dog	4	3	5	3	6	7	5	7	10	15	11
Aggressive span	2	5	2	6	8	4	4	17	6	10	11
Marked often	4	2	6	4	5	1	4	4	8	18	19
High strung	7	8	7	4	10	9	8	9	5	4	4
Good pet	3	3	4	6	5	5	6	7	9	6	21
Good service	15	17	11	5	8	3	2	6	2	1	5

The correlation between the rating “easy to train” and the degree of reinforcement needed was .766 ($p < .001$). The correlation is what one would expect, as dogs that respond well to human contact are likely to be more pleasant to train.

Novel situation. Within 3 weeks of training, the dogs were taken to PetSmart, a large local pet store for testing in a real world situation. The trainers tested dogs on their basic obedience (heel, sit, down, and stay) while inside the busy store. Dogs were rated on their confidence (defined as lack of fear and willingness to move about the store) and their ability to respond to commands. Thirty of the dogs were judged confident and responded easily to commands; 14 dogs were fearful, but still responded well to commands; 23 dogs were distracted and needed many corrections for correct responses; 5 dogs were very distracted and did not respond to commands even with corrections; 1 dog acted aggressive toward a male customer on the way into the store. It was decided that taking this dog into the store would not be safe. One dog had not reached criterion and was not tested in this situation.

Final retrieve tasks. As reported earlier, 40 dogs completed the retrieval portion of the training. These dogs' ability to retrieve was tested in three situations. The number of corrections, if any, needed for completion of the retrieve was recorded.

1. Retrieve in home Base situation: A dog should be expected to retrieve in the home base situation with few to no corrections. The mean number of corrections needed for retrieval in the home base situation was .625 (range of 0 to 8).

2. Retrieve with a male stranger present: A stranger can be threatening or distracting to some dogs, and it should be expected that more corrections would be needed in this situation, but that an ideal dog would not be affected. The mean

number of corrections needed for retrieval with a stranger present was 3.7 (range of 0 to 11).

3. Retrieve with loud noise: As in retrieve with a male stranger present, it is expected that the sound will be distracting but the dogs should be able to work quickly through the distraction. The mean number of corrections needed for retrieval with a loud noise was .9750 (range of 0 to 9).

Predictive Value of Test Items

The items on the selection test were designed to predict different traits, such as fear, dominance, and sensitivity to stimuli. The data indicate that several test items were able to predict certain traits. Furthermore, when a number of these test items are combined, a solid foundation for a final selection tool is apparent.

Fear. I define fear behavior as excessive sensitivity to stimuli, resulting in flight behavior. A dog who is to be used as a service animal or a performance dog should not be fearful. It is important that these dogs not be easily overwhelmed by novel stimuli. Two selection tests correlated highly with the trainer’s rating of fear. The Pearson correlation for sound sensitivity and the fear rating was 0.427 ($p < .001$). Figure 1 illustrates the strong relationship between these variables.

A dog who responded to the sound sensitivity test by initially startling to the sound and then approaching it was later rated by the trainers as less fearful than dogs who responded differently during selection. The dogs who startled and never approached the sound were rated the most fearful. The correlation between the fetch test and the fear rating was 0.377 ($p = .001$). In Figure 2, one can see that a dog who consistently interacts with the toy or the tester is less likely to be rated as fearful 5 weeks later by the trainers. A dog who does not interact with either the toy

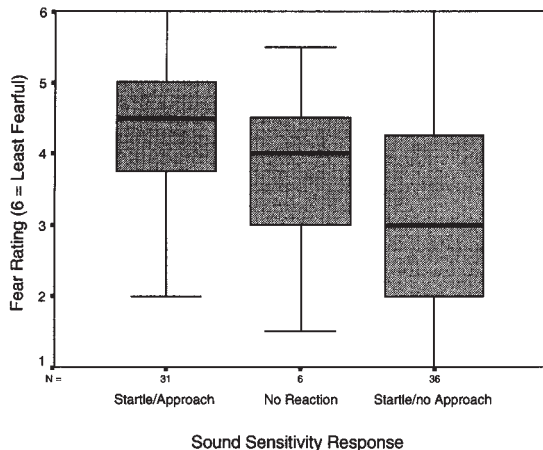


FIGURE 1 Relationship between sound sensitivity and the trainer’s rating of fearfulness (note that the high number on the y axis represents that the trainer believed that the dog displayed less of the trait. In most cases the high number also represents the desired behavior).

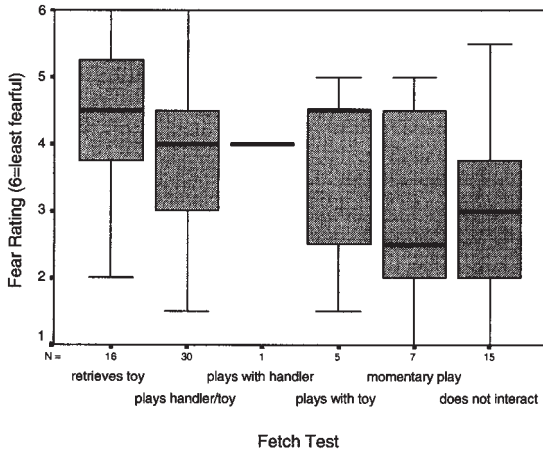


FIGURE 2 Relationship between fetch test and trainer's rating of fearfulness.

or the tester, or who interacts momentarily with the toy and then disengages from play, is more likely to be rated as fearful.

The opinion was also a good predictor of fear behavior. The Pearson R correlation coefficient for the variables was 0.333 ($p = .004$), indicating that some fear behavior can be predicted by the tester's opinions.

Submission. I define submission behavior as "effort on the part of an inferior or lower ranking animal toward creating friendly and harmonic social integration" (Overall, 1997). Consistent submissive behaviors such as body posture and ear posture help to determine this state. The pinch test proved to be a reliable measure of submission in the dogs. A dog who did not react at all when pinched was more likely to be rated as a submissive dog by the trainers after 5 weeks of training. Figure 3 shows this trend. The Pearson correlation between the pinch test and the trainer's rating of submission is 0.345 ($p = .003$).

The walk test also correlated with Submission, ($r = -0.229$, $p = .048$), with dogs who hung back and lay down during the selection test being rated as more submissive by the trainers 5 weeks later. Figure 4 displays the relationship between these variables. It is important to note that although only six dogs had the extreme response of creeping, these dogs were rated significantly more submissive than others.

High energy-high strung. The sensitivity test proved to be a good predictor of high-energy dogs. The dogs who did not stay still to accept the touch were more likely to be rated as high strung than were other dogs (see Figure 5). The Pearson correlation for these variables is .258 ($p = .026$). However, only eight dogs did not accept the touch during the sensitivity test, indicating that this test item might need further investigation.

The jumping on tester test also correlated with the high energy rating with a Pearson correlation ($r = 0.309, p = .007$). Dogs who jumped on the testers throughout the selection test process were more likely to be rated as being high strung than dogs that did not jump on the testers.

Dog aggression. The selection item reaction to other dog proved to be a good predictor of dog aggression observed during training with a Pearson correlation, $r = 0.341 (p = .004)$. Figure 6 shows the relationship between the variables. A dog that approached the “other dog” by rushing forward with its tail and head high was more likely to display dog aggressive behavior throughout training. The importance of determining dog aggression should not be overlooked for both service

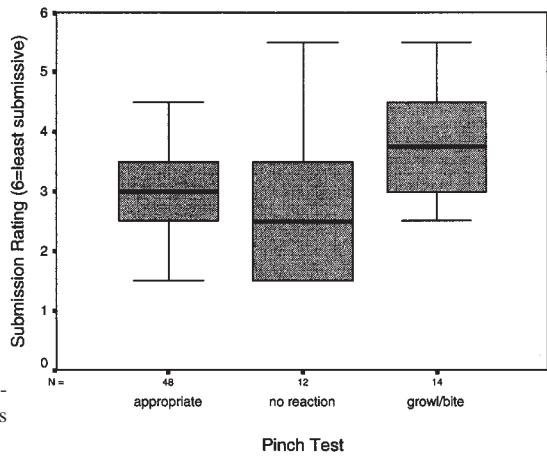


FIGURE 3 Relationship between pinch test and the trainer’s rating of submission.

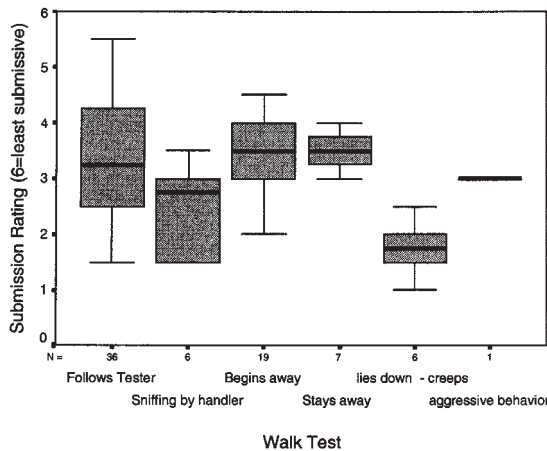


FIGURE 4 Relationship between the walk test and the trainer’s rating of submission.

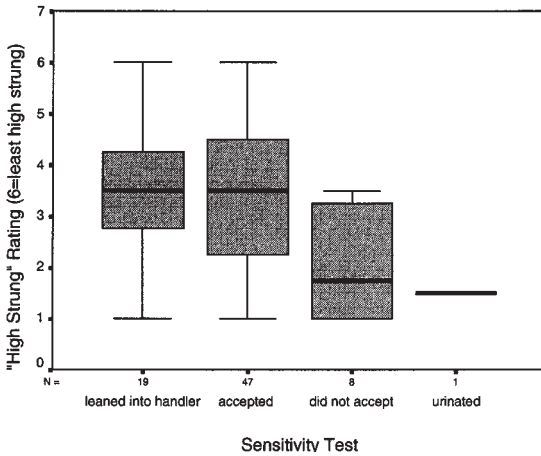


FIGURE 5 Relationship between the sensitivity test and the trainer's rating of the trait high strung.

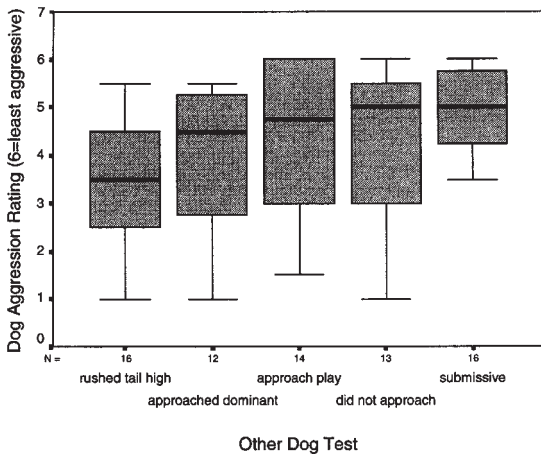


FIGURE 6 Relationship between the other dog test and the trainers' rating of dog aggressiveness.

dogs and pets. When determining whether a dog will fit in a particular environment, acceptance of other dogs should be an integral part of the decision.

Dominance. I define dominance as behaviors that lead to the acquisition of stimuli important to the individual dog, regardless of requests by other social members. Behaviors such as ear and body posture and “pushy” behaviors help determine this state. Two similar test items had a significant correlation with the trainers’ rating of dominance, activity level ($r = 0.264, p = .022$) and activity vertical ($r = 0.247, p = .033$). Dogs who crossed a higher number of squares were more likely to be rated as dominant than other dogs. Furthermore, dogs who raised onto their hind legs (activity level vertical) during the activity test were also more likely than other dogs to be rated as dominant.

The pinch test also correlated with dominance. The Pearson correlation, $r = -0.362$ ($p = .001$) when the reactions were ordered as no reaction, pulls back, and growl/bite. Figure 7 illustrates the relationship between these variables. A dog who growls or attempts to bite the tester is more likely to be rated as dominant by the trainers after 5 weeks of training.

Development of a Selection Tool

The development of a combination of selection items that best predict future behavior was one of the important aspects of this study. Because only 40 of 75 dogs learned the full retrieve task, only those 40 could be tested on the final tasks. A dog capable of not completing the retrieval task could still be rated high as a potential service dog. Unfortunately, monetary restraints governed the length of training time for each group. Because of this and the thoroughness of the trainers' ratings, it was decided that the trainer rating scale would be the primary determination of the dog's behavior during and at the completion of the 5 weeks of training. It was difficult, but possible, for a dog to not complete the retrieval task and still be deemed as having good service dog potential. The Pearson correlation between the trainer's assessment of the dog's potential for service work and whether the dog completed retrieval was 0.495 ($p < .001$).

To develop a scale that would encompass several important trainers' ratings, a reliability analysis was used. It was first thought that the trainer ratings of trainability, attention span, pet potential, and service dog potential should be used. However, pet potential, one of the factors, differs from service dog potential in several ways. A dog can be a good pet and be fearful of novel stimuli, or in some cases, aggressive toward other dogs. The correlation between the two ratings is 0.634 ($p < .001$). Although one would expect a high correlation, there are strong

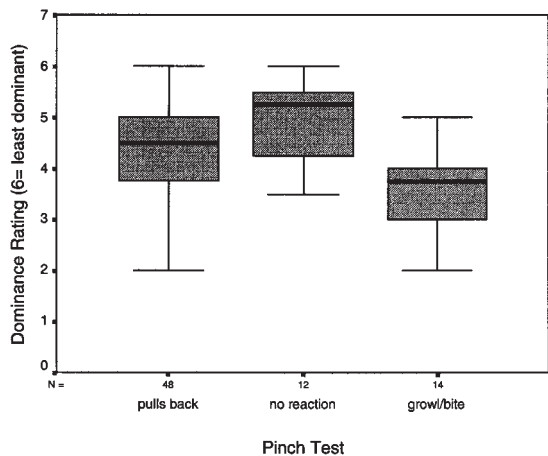


FIGURE 7 Relationship between the pinch test and the trainer's rating of dominance.

differences between a good pet and a good service dog. These differences are so strong that it is possible that leaving the pet rating in the scale might eradicate the effect of certain variables to predict important qualities for a service dog but not a pet. A pet can be dog aggressive or fearful of novel situations while a service dog cannot. Tolerance for these behaviors differs between pets and service dogs. The scale, "Service Success," was devised without the pet rating. It was comprised of trainability, service dog potential, and attention span. An $\alpha = .915$ is produced with this combination. A regression analysis correlating the service success variable to sensitivity, opinion, reaction to other dog, stranger, stare, and vertical activity level produced an $R = 0.603$ with an $F = 6.001$ ($p < 0.001$) accounting for 36.4% of the variance. Table 3 shows that several of the selection items are playing a strong role in the prediction of future behavior. Again the sensitivity test is weaker than the others, but because of its predictive value for individual traits, and the modest N , the test will remain for future examination.

DISCUSSION

Selection Tool to Date

I have been involved in the assessment of existing selection procedures and development of effective selection procedures for several years. In the course of this research, it was determined that the existing tests are not good predictors for selecting dogs from shelters for service training. This investigation was designed to begin the development of an effective selection tool. Sixteen items were investigated, with the anticipation that a few of the items would correlate with fu-

TABLE 3
Regression Analysis Service Success As the Dependent Variable

	<i>Coefficients</i>				
	<i>Unstandardized</i>		<i>Standardized</i>		<i>Zero-Order</i>
	<i>B</i>	<i>SE</i>	β	<i>p</i>	
Constant	4.017	3.062		.194	
Stranger	1.000	0.506	.206	.053	.275
Other dog	0.520	0.274	.196	.062	.134
Activity					
Vertical activity level	-0.445	0.190	-.238	.022	-.172
Stare	1.007	0.401	.261	.015	.345
Sensitivity	0.917	0.666	.151	.173	.331
Opinion	-0.542	0.195	-.299	.007	-.333

ture behavior. Not only did we find a solid base for the selection tool, but also several of the items correlated highly with specific traits that, in the past, had been difficult, if not impossible, to predict.

Sensitivity, opinion (or gut), reaction to other dog, stranger, stare, and vertical activity level were all selection items that are to be included in the selection tool. Combined, they account for over 36% of the variance. Two of these test items (sensitivity and vertical activity level) are novel to the field, and the others have been modified from existing selection tools.

Submission, fear, dog aggression, energy level, and dominance are all important traits when determining if a dog is appropriate for service work. Determination of these traits is also important for other canine disciplines in which selection plays a role in success. These include pounds and shelters, police work, search and rescue, and other canine tasks.

Those who are interested in determining a particular trait can take advantage of some of the test items involved in the tool. Thus, if one is interested in selecting a submissive dog, the pinch test and the walk test can be used to make the determination. The tool is designed to be easily administered by experienced professionals. It can be conducted in a relatively small period of time and can be administered before or after open hours at shelters and pounds.

The modifications to the selection items discussed previously should strengthen the selection tool, and future research will address those issues. Another factor that can improve the selection tool is the use of physiological measures that correlate with traits in canines. In the past several years, the use of cortisol levels and blood pressure have been investigated as predictors of behavior in canines (Beerda, Schilder, Janssen, & Mol, 1996; Beerda, Schilder, van Hooff, & de Vries, 1997; Vincent & Michell, 1992, 1996). The use of cortisol has been made more accessible by research that has determined saliva to be a good source of accurate cortisol data (Vincent & Michell, 1992).

THE REGRESSION ANALYSIS

The regression analysis accounted for over 36% of the variance, signifying that the selection tool is developing well. By accounting for over a third of the variability, the tool should help decrease the number of dropouts. The selection items that are incorporated in the regression are strong predictors of certain aspects of the dogs' behavior. Vertical activity level apparently is a completely new selection item as the author does not know of any service dog agency that uses such an item. The sensitivity item also is a new item for selection. The items are all relatively easy to administer and can be standardized without any difficulty.

The opinion is the only item that is dependent on a thorough familiarity and experience with dogs. All selection testers involved in the study had at least 5 years

of experience, the average being 10 years. It is important for service dog centers to employ a tester with good working experience with dogs.

The vertical activity level test is both simple and completely objective. These qualities add to its strength. The stare test is not as easy to control as the activity test, but it is a simple straightforward test to administer and produced good results. It is possible that one of the reasons it was a strong predictor was because it was always the first test item. The dogs did not have the opportunity to establish a relationship with the tester that might have changed the dogs' response.

Although the sensitivity test did not have as large of a β weight as some of the other items in the regression analysis, it did have some effect on the analysis. I believe that its prediction strength for particular traits makes it a worthwhile test for future research. This too is true of the stranger test, which can be strengthened by consistently testing this item late in the selection process, when the dogs have had the opportunity to settle and relax.

The other dog test is unique as it predicts dog aggression without producing overt aggression. Dog aggression is a serious problem for a potential service dog to have as this trait often causes a dog to be virtually unable to concentrate and work when other dogs are present.

In conclusion, this study indicates that it is possible to improve the selection process for service animals. Improved selection will decrease the amount of drop-outs, therefore increasing the number of people able to benefit from these animals. The benefits include richer lives for those with disabilities: decreasing health care costs, decrease in the need for health care professionals to assist in daily life, fewer dogs in shelters and pounds, and improved human–animal interactions.

REFERENCES

- Bartlett, M. (1979, March). A novice looks at puppy testing. *Purebred Dogs/American Kennel Gazette*, 31–42.
- Bartlett, M. (1985, March). Puppy aptitude testing: A new look. *Purebred Dogs/American Kennel Gazette*, 30–35, 64.
- Beaudet, R., Chalifoux, A., & Dallaire, A. (1994). Predictive value of activity level and behavioral evaluation on future dominance in puppies. *Applied Animal Behavior Science*, 40, 273–284.
- Beerda, B., Schilder, M. B. J., Janssen, N., & Mol, J. A. (1996). The use of saliva cortisol, urinary cortisol and catecholamine measurements for a noninvasive assessment of stress responses in dogs. *Hormones and Behavior*, 30, 272–279.
- Beerda, B., Schilder, M. B. H., van Hooff, J., & de Vries, H. W. (1997). Manifestations of chronic and acute stress in dogs. *Applied Animal Behaviour Science*, 52, 307–319.
- Blasch, B. B., & Stuckey, K. A. (1995, September/October). Accessibility and mobility of persons who are visually impaired: A historical analysis. *Journal of Visual Impairment and Blindness*, 417–422.
- Campbell, W. E. (1972). A behavior test for puppy selection. *Modern Veterinary Practice*, 12, 29–33.
- Goddard, M. E., & Beilharz, R. G. (1983). A multivariate analysis of the genetics of fearfulness in potential guide dogs. *Behavior Genetics*, 15, 69–89.

- Goddard, M. E., & Beilharz, R. G. (1984). The relationship of fearfulness of sex, age, and experience on exploration and activity in dogs. *Applied Animal Behaviour Science*, *12*, 267–278.
- Goddard, M. E., & Beilharz, R. G. (1985a). Early prediction of adult behavior in potential guide dogs. *Applied Animal Behaviour Science*, *15*, 247–260.
- Goddard, M. E., & Beilharz, R. G. (1985b). Individual variation in agonistic behaviour in dogs. *Animal Behaviour*, *33*, 1338–1342.
- Goddard, M. E., & Beilharz, R. G. (1986). Early prediction of adult behaviour in potential guide dogs. *Applied Animal Behaviour Science*, *15*, 247–260.
- Martinek, Z. (1972). The process of habituation as a test of interindividual (typological) differences in behavior of dogs. *Acta Neurobiology Experimental*, *33*, 791–801.
- Overall, K. L. (1997). *Clinical behavioral medicine for small animals*. St. Louis, MO: Mosby Year Book.
- Pfaffenberger, C. J., Scott, J. P., Fuller, J. L., Ginsburg, B. E., & Bielfelt, S. W. (1976). *Guide dogs for the blind: Their selection, development, and training*. Amsterdam: Elsevier Scientific.
- Prestrude, A., & O'Shea, J. (1998). Dogs in service to humans. In G. Greenberg & M. Haraway (Eds.), *Comparative psychology: A handbook* (pp. 386–392). New York: Garland.
- Scott, J. P., & Fuller, J. L. (1965). *Genetics and the social behavior of the dog*. Chicago: University of Chicago Press.
- Van der Borg, J., Netto, W., & Planta, D. (1991). Behavioral testing of dogs in animal shelters to predict problem behavior. *Applied Animal Behaviour Science* *32*, 237–251.
- Vincent, I. C., & Michell, A. R. (1992). Comparison of cortisol concentrations in saliva and plasma of dogs. *Research in Veterinary Science*, *53*, 342–345.
- Vincent, I. C., & Michell, A. R. (1996). Relationship between blood pressure and stress prone temperament in dogs. *Physiology & Behavior*, *60*, 135–138.
- Volhard, J., & Volhard, W. (1992). *Open and utility training: The motivational method*. New York: MacMillan.
- Weiss, E., & Greenberg, G. (1997). An evaluation of service dog selection tests: Effectiveness for dogs from animal shelters. *Applied Animal Behaviour Science*, *53*, 297–308.
- Weiss, E., & Greenberg, G. (1998). Open field procedures. In G. Greenberg & M. Haraway (Eds.), *Comparative psychology: A handbook* (pp. 257–263). New York: Garland.
- Wilsson, E., & Sundgren, E. (1998). Behaviour test for eight-week old puppies—Heritabilities of tested behaviour traits and its correspondence to later behaviour. *Applied Animal Behaviour Science*, *58*, 151–162.

APPENDIX
Trainer's Assessment Questionnaire

1. This dog was easy to train.
Agree 6 5 4 3 2 1 Disagree
2. This dog was fearful.
Agree 6 5 4 3 2 1 Disagree
3. This dog was dominant.
Agree 6 5 4 3 2 1 Disagree
4. This dog was submissive.
Agree 6 5 4 3 2 1 Disagree
5. This dog was dog aggressive.
Agree 6 5 4 3 2 1 Disagree
6. This dog had a good attention span.
Agree 6 5 4 3 2 1 Disagree
7. This dog marked often.
8. This dog was high strung.
Agree 6 5 4 3 2 1 Disagree
9. This dog would make a good pet.
Agree 6 5 4 3 2 1 Disagree
10. This dog would make a good service dog.
Agree 6 5 4 3 2 1 Disagree